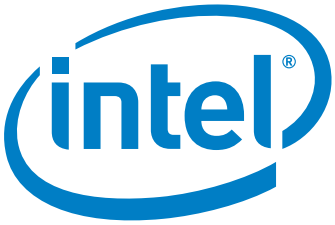
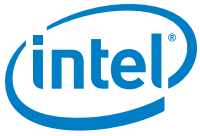


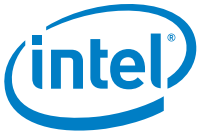
Intel® Omni-Path Fabric Software in SUSE* Linux* Enterprise Server 12 SP3

Release Notes

November 2019

Order No.: J71758-4.0





Contents

Tables

§

1.0 Overview of the Release

1.1 Introduction

This document provides a brief overview of the changes introduced into the Intel® Omni-Path Software by this release. References to more detailed information are provided where necessary. The information contained in this document is intended as supplemental information only; it should be used in conjunction with the documentation provided for each component.

These Release Notes list the features supported in this software release, open issues, and issues that were resolved during release development.

1.2 Audience

The information provided in this document is intended for installers, software support engineers, service personnel, and system administrators.

1.3 Software License Agreement

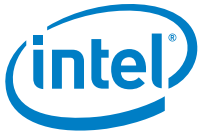
This software is incorporated into the SUSE* Linux* Enterprise Server* (SLES*) 12 SP3 distribution and is covered under SLES* 12 SP3 licensing.

1.4 If You Need Help

Technical support is available from SUSE* support.

1.5 Packages in this Release

Intel® Omni-Path Software Packages
Packages created by Intel
opa-basic-tools-10.3.1-2.5.x86_64.rpm
opa-fm-10.3.1-4.15.x86_64.rpm
opa-fastfabric-10.3.1-2.5.x86_64.rpm
opa-fmgui-10.1.0.0-1.33.noarch.rpm
libpsm2-2-10.2.103-2.6.x86_64.rpm
libpsm2-devel-10.2.103-2.6.x86_64.rpm
libpsm2-compat-10.2.103-2.6.x86_64.rpm
Firmware binaries delivered by Intel
8051 firmware version 1.18



Intel® Omni-Path Software Packages (Continued)
SBus Master firmware version 0x10130001
PCIe SerDes firmware version 0x4755
Fabric SerDes firmware version 0x1055
Packages used by Intel
rdma-core-14-6.7.x86_64
openmpi-1.10.6-2.19.x86_64
mpitests-openmpi-3.2-8.18.x86_64
mvapich2-psm-2.2-12.6.x86_64
mpitests-mvapich2-psm-3.2-8.19.x86_64

1.6 Supported Features

- The list of supported hardware is in Table 1-2.
- Coexistence with Intel® True Scale Architecture. This release supports True Scale hardware serving as an InfiniBand* storage network with the Intel® Omni-Path hardware used for computing. Note that connecting a True Scale adapter card to an Omni-Path switch, or vice-versa, is not supported. For more details on this feature, refer to *Intel® Omni-Path Fabric Host Software User Guide*.
- Supports Dual Rail: Two Intel® Omni-Path Host Fabric Interface (HFI) cards in the same server connected to the same fabric
- Supports Dual Plane: Two HFI cards in the same server connected to separate fabrics.
- Limited validation testing performed on network storage file systems:
  - NFS over TCP/IP
- Active Optical Cables. For details, see the Cable Matrix at: http://www.intel.com/content/www/us/en/high-performance-computing-fabrics/omni-path-cables.html
- MPI applications are provided in a stand-alone package.
- Monitored Intel® Omni-Path Host Fabric Interface
- DHCP and LDAP supported on Intel® Omni-Path Edge Switch 100 Series and Intel® Omni-Path Director Class Switch 100 Series hardware.
- Support for the Enhanced Hypercube Routing Engine is outside the scope of Intel® OPA support. However, Intel partners may offer such support as part of their solutions. In addition there is an open source community who may be able to answer specific questions and provide guidance with respect to the Enhanced Hypercube Routing Engine.

1.7 Supported MPI Libraries

The table below lists the different MPI libraries supported by Intel® Omni-Path Software. Note that the second column indicates whether the MPI library is included in the distribution or not.

Table 1-1. Supported MPI Libraries

MPI Implementation	Included in distribution?	Runs Over
Open MPI 1.10.6	Yes	PSM2
MVAPICH2-2.2	Yes	PSM2

Note: Refer to the *Intel® Omni-Path Fabric Host Software User Guide* for set up information when using Open MPI with the SLURM PMI launcher and PSM2.

1.8 Intel Hardware

Table 1-2 lists the Intel hardware supported in this release.

Note: The Intel® PSM2 implementation has a limit of four (4) HFIs.

Table 1-2. Supported Hardware

Hardware	Description
Intel® Xeon® Processor E5-2600 v3 product family	Haswell CPU-based servers
Intel® Xeon® Processor E5-2600 v4 product family	Broadwell CPU-based servers
Intel® Xeon® Scalable Processor	Skylake CPU-based servers
Intel® Xeon Phi™ Processor x200 product family	Knights Landing CPU-based servers
Intel® Omni-Path Host Fabric Interface 100HFA016 (x16)	Single Port Host Fabric Interface (HFI)
Intel® Omni-Path Host Fabric Interface 100HFA018 (x8)	Single Port Host Fabric Interface (HFI)
Intel® Omni-Path Switch 100SWE48Q	Managed 48-port Edge Switch
Intel® Omni-Path Switch 100SWE48U	Externally-managed 48-port Edge Switch
Intel® Omni-Path Switch 100SWE48UFH	Externally-managed 48-port Edge Switch, hot-swap power and fans
Intel® Omni-Path Switch 100SWE48QFH	Managed 48-port Edge Switch, hot-swap power and fans
Intel® Omni-Path Switch 100SWE24Q	Managed 24-port Edge Switch
Intel® Omni-Path Switch 100SWE24U	Externally-managed 24-port Edge Switch
Intel® Omni-Path Director Class Switch 100SWD24	Director Class Switch 100 Series, up to 768 ports
Intel® Omni-Path Director Class Switch 100SWD06	Director Class Switch 100 Series, up to 192 ports

1.9 Intel® OPA Compatibility Matrix

The following component versions are compatible with Intel® Omni-Path software in SLES* 12 SP3.

Table 1-3. Intel® OPA Compatibility Matrix

UEFI	TMM	Managed Switch	Externally-Managed Switch	FM GUI
1.4.2.0.0	10.4.0.0.146	10.4.3.0.1	10.4.3.0.1	10.4.0.0.184
1.3.4.0.0	10.2.1.0.3	10.3.1.0.8	10.3.1.0.8	10.3.0.0.60
0x29	10.2.0.0.154	10.2.0.0.154	10.2.0.0.152	10.2.0.0.152

1.10 Installation Instructions

Perform the steps in this section to install the default Intel® Omni-Path Software configuration.

Assumptions

- You are logged in as root or with root privileges.
- You have a list of IPv4 addresses and netmasks for each IPoIB interface you are going to set up.
- SLES* packages are available in a zypper repository.

References

- Refer to the *Intel® Omni-Path Fabric Software Installation Guide* for related software requirements and next steps.
- Refer to the *Intel® Omni-Path Fabric Hardware Installation Guide* for related firmware requirements.

Table 1-4. Installation Instructions (Sheet 1 of 2)

Step	Task/Prompt	Action
On all nodes in the fabric: Install OPA-Basic Software		
1.	At the command prompt, enter the installation command for `opa-basic-tools`.	Type `zypper install -y opa-basic-tools` and press **Enter**.
2.	At the command prompt, reboot the server.	Type `reboot` and press **Enter**.

3.	Check your link using the `opainfo` command. PortState shows Init (LinkUp).	Type `opainfo` and press **Enter**.
		Example output:
		`hfi1_0:1`
		`PortGUID:0x001175010163cf87`
		`PortState: Init (LinkUp)`
		`LinkSpeed Act: 25Gb En: 25Gb`
		`LinkWidth Act: 4 En: 4`
		`LinkWidthDnGrd ActTx: 4 Rx: 4 En: 1,2,3,4`
		`LCRC Act: 14-bit En: 14-bit,16-bit,48-bit`
		`QSFP: PassiveCu, 1m FCI Electronics P/N 10131941-2005LF Rev 6`
		`Xmit Data: 2 MB Pkts: 21761`
		`Recv Data: 10 MB Pkts: 21761`
		`Link Quality: 5 (Excellent)`

4.	Install the `rdma-core` rpm.	Type `zypper install -y rdma-core` and press **Enter**.
5.	On all compute nodes: install the PSM2 library.	Type `zypper install -y libpsm2-2` and press **Enter**.

Table 1-4. Installation Instructions (Sheet 2 of 2)

Step	Task/Prompt	Action
6.	(Optional) On one node in the fabric: install the Fabric Manager GUI.	Type `zypper install -y opa-fmgui` and press **Enter**.
On the management node: Install Intel® Omni-Path Fabric Suite Components		
7.	Install FastFabric.	Type `zypper install -y opa-fastfabric` and press **Enter**.
8.	Install Fabric Manager.	Type `zypper install -y opa-fm` and press **Enter**.
9.	Start the Fabric Manager.	Type `service opafm start` and press **Enter**.
10.	Install the PSM2 library, if not already installed.	Type `zypper install -y libpsm2-2` and press **Enter**.
On all nodes in the fabric: Set up IPoIB IPV4 Configuration		
11.	Manually edit or create the `ifcfg-ib`*X* file.	***Note:*** Use the OS distribution-supplied instructions for setting up network interfaces.

		Type `cat /etc/sysconfig/network/ifcfg-ib0` and press **Enter**.
		Example output:
		`DEVICE=ib0`
		`BOOTPROTO=static`
		`IPADDR=10.228.200.173`
		`BROADCAST=10.228.203.255`
		`NETWORK=10.228.200.0`
		`NETMASK=255.255.252.0`
		`ONBOOT=yes`
		`CONNECTED_MODE=yes`
		`MTU=65520`

12.	Bring up the ib0 interface.	Type `ifup ib0` and press **Enter**.

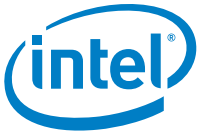
13.	Check your link using the `opainfo` command. PortState shows Active.	Type `opainfo` and press **Enter**.
		Example output:
		`hfi1_0:1`
		`PortGID: 0xfe80000000000000:001175010163f931`
		`PortState: Active`
		`LinkSpeed Act: 25Gb En: 25Gb`
		`LinkWidth Act: 4 En: 4`
		`LinkWidthDnGrd ActTx: 4 Rx: 4 En: 3,4`
		`LCRC Act: 14-bit En: 14-bit,16-bit, 48-bit`
		`Mgmt: True`
		`LID: 0x00000010-0x00000010 SM LID: 0x0000000c`
		`SL: 0`
		`QSFP: AOC,5m FINISAR CORP P/N FCBN425QB1C05 Rev A`
		`Xmit Data: 0 MB Pkts: 251`
		`Recv Data: 0 MB Pkts: 251`
		`Link Quality: 5 (Excellent)`

14.	Perform a test ping.	Type `ping <remote IPoIB IP address>` and press **Enter**.
		For example:
		`ping 10.228.200.161`
		`PING 10.228.200.161 (10.228.200.161) 56(84) bytes of data.`
		`64 bytes from 10.228.200.161: icmp_seq=1 ttl=64 time=0.863 ms`

End task.		

1.11 Product Constraints

None.



1.12 Product Limitations

- PA Failover should **not** be enabled with FMs running on differing software versions. To disable PA failover, edit the `/etc/sysconfig/opafm.xml` file and in the `<Pm>` section, change `<ImageUpdateInterval>` to `0`.
- Enabling UEFI Optimized Boot on some platforms can prevent the HFI UEFI driver from loading during boot. To prevent this, do not enable UEFI Optimized Boot.

1.13 Documentation Versions

Table 1-5 lists the end user document versions supported by this release.

Table 1-5. Supported Documentation Versions

Title	Doc. Number	Revision
Key: Shading indicates the URL to use for accessing the particular document.		
• Intel® Omni-Path Switches Installation, User, and Reference Guides: http://www.intel.com/omnipath/SwitchPublications		
• Intel® Omni-Path Software Installation, User, and Reference Guides (includes HFI documents): http://www.intel.com/omnipath/FabricSoftwarePublications		
• Drivers and Software (including Release Notes): http://www.intel.com/omnipath/Downloads		
Intel® Omni-Path Fabric Setup Guide	J27600	4.0
Intel® Omni-Path Fabric Switches Hardware Installation Guide	H76456	5.0
Intel® Omni-Path Host Fabric Interface Installation Guide	H76466	3.0
Intel® Omni-Path Fabric Software Installation Guide	H76467	5.0
Intel® Omni-Path Fabric Switches GUI User Guide	H76457	5.0
Intel® Omni-Path Fabric Switches Command Line Interface Reference Guide	H76458	5.0
Intel® Omni-Path Fabric Suite FastFabric User Guide	H76469	5.0
Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide	H76472	5.0
Intel® Omni-Path Fabric Suite Fabric Manager User Guide	H76468	5.0
Intel® Omni-Path Fabric Suite Fabric Manager GUI User Guide	H76471	5.0
Intel® Omni-Path Fabric Host Software User Guide	H76470	5.0
Intel® Performance Scaled Messaging 2 (PSM2) Programmer's Guide	H76473	5.0
Intel® Omni-Path Fabric Performance Tuning User Guide	H93143	6.0
Intel® Omni-Path IP and Storage Router Design Guide	H99668	4.0
Building Lustre Servers with Intel® Omni-Path Architecture Application Note*	J10040	1.0
Building Containers for Intel® Omni-Path Fabrics using Docker and Singularity* Application Note*	J57474	1.0
Intel® Omni-Path Fabric Software Release Notes	J52019	1.0
Intel® Omni-Path Fabric Fabric Manager GUI Release Notes	J44214	1.0
Intel® Omni-Path Fabric Switches Release Notes (includes managed and externally-managed switches)	J36083	1.0

2.0 Issues

2.1 Introduction

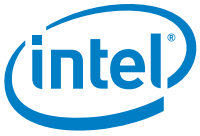
This section provides a list of the resolved and open issues in the Intel® Omni-Path Software.

2.2 Resolved Issues

Table 2-1 lists issues that are resolved in this release.

Table 2-1. Issues resolved in this release

ID	Description	Resolved in Release
132219	Server platforms running IFS 10.3.0 release (or Intel® OPA software delivered in certain Linux* OS distributions) and using integrated HFI for OPA (commonly known as "-F") may not support Active Optical Cables (AOC) after boot up.	SLES* 12SP3
133377	irqbalance settings are not being honored correctly after a reboot.	SLES* 12SP3
134124	HFI port stuck in INIT state due to SM failure to set pkeys.	SLES* 12SP3
134135 134429	When running communication-intensive workloads with 10KB MTU, it is possible to encounter node and/or job failures.	SLES* 12SP3
134471	The HFI UEFI driver cannot boot via PXE using Grub 2.	SLES* 12SP3
134772	`opatmmtool` will fail if provided with a filename (full path) that is longer than 63 characters.	SLES* 12SP3
134956	ib0 fails to become ready on warm reboots.	SLES* 12SP3
135000	Fabric Manager configuration files that specify IncludeGroup fields with undefined or nonexistent device groups could cause Fabric Manager failure.	SLES* 12SP3
135729 135870	KNL-F/SKL-F ports are offline in pre-boot setting when connected with AOC.	SLES* 12SP3
135812	FM may crash and restart in the event of a failure during topology assignments. This may result in mismatched port physical states on a link. While unlikely, this event may occur when there are integrity issues on a link.	SLES* 12SP3
135958	Spurious segmentation faults with greater than 2MB PSM2 transfers on Intel® Xeon Phi™ platforms.	SLES* 12SP3
136028	Two versions of the UEFI firmware are contained in the hfi-uefi RPM in the 10.3.0 IFS and BASIC packages. The files are functionally identical except the unsigned files (`HfiPcieGen3Loader_<version number>.unsigned.rom` and `HfiPcieGen3_<version number>.unsigned.efi`) are not signed for secure boot.	SLES* 12SP3
136152	Server platforms using integrated HFI for OPA (commonly known as "-F") require BIOS that provides UEFI version 1.3.1.0.0 and a configuration data file for pre-boot support of Active Optical Cables (AOC). Some servers may not have these files available in BIOS and will therefore not support AOC in pre-boot.	SLES* 12SP3
136318	SM crashes showing segfault errors in logs and high CPU usage. These crashes were caused by a mismatch of pahistory file versions.	SLES* 12SP3
136621	PCIe Fatal Errors during reboot cycles on server platforms using integrated HFI for OPA (commonly known as "-F").	SLES* 12SP3



2.3 Open Issues

Table 2-2 lists the open issues for this release.

Table 2-2. Open Issues (Sheet 1 of 2)

ID	Description	Workaround
131745	When running OpenMPI 1.10.0 on SLES* 12 with large number of ranks per node (over 40), it may happen that the ORTE daemon (orted) "hangs" during the finalization of job.	Stopping and resuming the "hung" orted process allows the job to finish normally. To find the hung process, run the ps and find a node with several job zombie processes. In that same node, identify the orted process ID and send a stop signal (kill -19 <PID>) and a continue signal (kill -18 <PID>).
134353	Very infrequently, when a link goes down, the logical link state can remain stuck in the 'Init' state.	The device containing the affected port must be rebooted in order to resolve the issue. Ports in this state have a logical link state of 'Init' but do NOT have a physical port state of 'LinkUp'.
134493	When using Mvapich2 with Intel® Omni-Path PSM2, users will notice unexpected behavior when seeding the built-in random number generator with functions like srand or srandom before MPI_Init is called. MPI_Init re-seeds the random number generator with its own value and does not restore the seed set by the user application. This causes different MPI ranks to generate different sequences of random numbers even though they started with the same seed value.	Seed the random number generator after MPI_Init is called or use the reentrant random number generator functions such as drand48_r.
135040	You can't currently specify portions of an Intel® Director Class Switch chassis that is not populated and is not expected to be populated. If CoreFull is 1, all the internal links for that chassis are generated when run against `opaxlattopology`. If CoreFull is 0, none of the links are generated.	Copy internal configuration of desired Intel® Director Class Switch into the main topology tab of the spreadsheet. Then delete all lines corresponding to leafs or spines that are not present in the configuration.
135259	On rare occasions, the HFI links do not come up after a reboot.	Reboot or bounce the link.
135326	Calling `opasmaquery` fails when called from a non-SM node to a node which has not booted to the OS.	Use the SM node when calling `opasmaquery` in this way.
135929	Intel® Omni-Path Boot nodes occasionally dropped from fabric when switching master SM from one node to another.	Reboot PXE client node.
135951	When creating host verify punchlist, the following error message is displayed: `unable to parse filter -s Invalid slot number`	None.
135975	After performing an OPA software configuration update, some unmanaged switches do not update the settings for LinkWidth and LinkWidthDnGrade enables.	A reboot is required for configuration changes made to an externally managed switch to become active.
136160	On some Intel® Xeon Phi™ with integrated Intel® Omni-Path fabric platforms, the second integrated HFI is discovered first and is subsequently identified as the first HFI device. As a result, when issuing Intel® Omni-Path commands, the second HFI appears first in the results. In Linux* and various Intel® Omni-Path tools, the HFI reporting order may be the opposite of the order appearing on the Intel® Xeon Phi™ with integrated Intel® Omni-Path fabric cable/faceplate.	You can identify the second integrated HFI by inspecting the Node GUID or Port GUID/Port GID reported by `opainfo` or other commands such as `hfi1_control -i`. Note that bit 39 of the PortGUID, the most significant bit, is set for the second HFI, and is clear for the first HFI. Keep in mind that when issuing various Intel® Omni-Path CLI commands targeted at a specific HFI using the `-h` option, `-h 1` correlates to the device that is listed as `hfi1_0`. As a result, the issued command affects the second HFI instance in cases where the second HFI port instance appears first.

Table 2-2. Open Issues (Sheet 2 of 2)

ID	Description	Workaround
136728	If hundreds of links are bouncing while the FM is sweeping, the FM sweep time may be significantly extended. This can result in unexpected delays in FM responsiveness to fabric changes or host reboots. (The issue is that active links bounce between the time FM discovers one side of the link versus the other side of the link.) In this release a fix was included to address the situation that occurs in fabrics of >1000 nodes when numerous links bounce (or hosts are rebooted) at once.	The following workarounds are recommended: • When rebooting nodes on a production cluster, perform reboots in batches of 300 nodes or less. • During cluster deployment, carefully follow the procedures in the *Intel® Omni-Path Fabric Staging Guide* and use FastFabric to check signal integrity and placement of all cables. Correct or disable any problematic links before starting production use of the cluster. • When replacing or expanding a production cluster, repeat the procedures in the *Intel® Omni-Path Fabric Staging Guide* to verify the new hardware. Correct or disable any problematic links before resuming production use of the cluster. • Use the PM, FM logs, FM GUI, FastFabric, and other tools to monitor signal integrity and link stability. Correct or disable any problematic links when discovered.
136733	Slow memory deregistration has been observed.	None.
136822	AOC support is not available on integrated HFI platforms (-F platforms) if the Intel UEFI driver is not executed during boot. Some BIOS will not execute the UEFI driver in Legacy BIOS boot mode. Also, some BIOS configuration settings or other system settings will bypass execution of the HFI driver.	Avoid the use of Legacy BIOS boot mode if your platform does not execute the HFI driver in that mode. Avoid BIOS settings or other configuration settings that do not execute the HFI driver during boot.
136901	Occasionally, pre-boot nodes are dropped by the Fabric Manager during fabric sweeps, where the system containing the dropped pre-boot node has more than one HFI on a single socket.	Bounce the link of the dropped pre-boot port.

136902	A snapshot file with a multicast group with rate 10g will not be read properly.
	The following error is returned:
	opafabricanalysis: Port 0:0 Error: Unable to analyze fabric snapshot. See /var/usr/lib/opa/analysis/latest/fabric.0:0.links.stderr
	opafabricanalysis: Possible fabric errors or changes found
		On all nodes running opafm, run:
		`systemctl stop opafm`
		On all switches running ESM, run:
		`smControl stop`
		For all nodes/servers running ibacm:
		1. Create a file /etc/rdma/ibacm_opts.cfg with one line:
		`min_rate 40`
		2. Restart ibacm
		
		In the nodes running the host FM, restart opafm or start opafm with the command:
		`systemctl start opafm`
		In the switches running ESM, run:
		`smControl start`

136985	opahfirev has output errors when the HFI driver is not installed.	None.
136995	The opahfirev tool output uses the term “HWRev” to indicate the revision of the silicon on the card.	None.
137054	Pinging an Intel® OPA UEFI permanent IP address from a DHCP server fails on subsequent reboots unless the corresponding network interface has first been initialized in the UEFI network stack.	Before pinging a UEFI permanent IP address, first initialize the corresponding network interface in the UEFI network stack.
137221	Querying for switch info with opasmaquery while using the -g option will print incorrect IPv4 addresses.	Do not use the -g option.

§ §